

• 综述 •

基于全基因组关联研究的乳腺癌预后预测研究现状

叶娟^{1,4} 郭翠萍² 张国君^{1,3}

【摘要】 全基因组关联研究可用于确认与某些复杂疾病相关的单核苷酸多态性位点,目前运用该方法已鉴定出多个乳腺癌易感基因位点,通过该方法对高风险位点的研究或许能为乳腺癌预后的预测提供重要信息。笔者总结了基于全基因组关联研究的乳腺癌预后预测的研究现状,以期能为乳腺癌的治疗提供参考。

【关键词】 乳腺肿瘤; 全基因组关联研究; 预后

【中图分类号】 R737.9; R596.3

【文献标志码】 A

2012年全球癌症数据库 GLOBOCAN 的数据显示,乳腺癌为全球女性发病率最高的恶性肿瘤,新发病例每年约 167 万,死亡病例约 52 万^[1]。随着乳腺 X 线筛查的推广及治疗水平的提高,乳腺癌患者的生存率有所上升^[2],但不同患者对治疗的敏感性及其预后经常难以预测。乳腺癌患者的预后受多因素影响,包括年龄、肿瘤大小、肿瘤的组织学分级、诊断时局部及远处转移的程度、激素受体状态、HER-2 表达情况以及患者所接受的治疗、对治疗的依从性等^[3]。除传统的临床病理特征外,肿瘤基因组谱系表达研究表明癌症分子信号也许能够提供更准确的预后信息,但该方法目前尚未在临床上常规应用。遗传学相关研究提示遗传因素也对乳腺癌预后具有重要意义^[4]。早期的候选基因关联研究策略主要分析与乳腺癌发生、发展相关的基因位点,如 DNA 修复过程、雌激素代谢、炎症反应、血管生成和细胞周期调节等^[5]。尽管这些研究发现的单核苷酸多态性(single nucleotide polymorphism, SNP)位点与乳腺癌预后存在一定的相关性,但大量的研究仍存在样本量不足、结果不一致、无法重复等缺点。近年来,全基因组关联研究(genome-wide association study, GWAS)以其大样本量、基因检测覆盖面广等优势已成功发现并鉴定了与乳腺癌遗传易感性相关的多个 SNP 位点^[6],这也提示 GWAS 在研究乳腺癌预后方面具有一定的潜在价值。笔者总结了基于 GWAS 的乳腺癌预后预测的研究现状,并探讨其潜在的应用价值。

一、GWAS

GWAS 是在人类全基因组范围内对不同个体的遗传变异进行基因分型,计算序列变异与疾病或表型的关联程度,选择差异具有统计学意义的 SNP 并进行验证,最终确认与疾病或表型关联的 SNP^[7]。GWAS 在全基因组范围内进行

研究,不再需要像候选基因策略一样预先构建假设,即不需在研究之前假设某些基因或位点与疾病的相关性,可以在一定程度上避免偏倚。2005 年,Edwards 等^[8]报道了年龄相关性视网膜黄斑变性的 GWAS,之后 GWAS 广泛应用于研究各种复杂疾病,如肥胖^[9]、肿瘤^[10]、2 型糖尿病^[11]、精神类疾病^[12]、心血管疾病^[13]等。然而, GWAS 研究的工作量大,费用昂贵,开展研究需同时具备以下条件:合乎条件的研究样本,即病例组样本须携带致病的遗传因素;足够的检验效能,满足这一条件需要足够数目的样本量和 SNP;合适的统计分析方法^[14]。尽管目前利用 GWAS 已经发现很多可能与某些复杂疾病相关的 SNP^[15],但这些结果尚未能在临床应用,基因与基因之间以及基因与环境之间如何相互作用导致疾病发生还有待进一步研究。

二、乳腺癌易感性的 GWAS 进展

2007 年,乳腺癌的 GWAS 被报道,目前已有超过 25 项 GWAS 报道与乳腺癌相关的 SNP,这些易感位点达 190 个^[16]。Stacey 等^[17]对来自冰岛的家系性乳腺癌患者和健康对照检测了约 300 000 个 SNP,结果发现 2 个易感位点 rs3803662、rs13387042,分别位于 16q12 和 2q35,并且与 ER 阳性的乳腺癌关联性更强。Turnbull 等^[18]在英国人群中开展的 GWAS 同样也纳入了有乳腺癌家族史的患者,发现了 5 个新的遗传易感位点(rs1011970、rs2380205、rs10995190、rs704010、rs614367),分别位于 9、10、11 号染色体,同时还验证了其他 SNP 与患癌风险增加相关,包括 rs2981582(FGFR2)、rs13281615(8q24)、rs3817198(LSP1)等。Zheng 等^[19]在中国人群中进行的 GWAS 发现:位于 6q25.1 的 ESR1 基因上游的 rs2046210 是重要的乳腺癌易感位点。Long 等^[20]在另一项亚洲人群的 GWAS 中也发现了另一个与乳腺癌相关的 SNP(rs9485372),位于 6q25.1。Cai 等^[21]的研究提示 rs10822013(10q21.2)有可能是中国女性特有的乳腺癌易感基因位点。

三、关于乳腺癌预后的高风险位点的 GWAS 研究

Easton 等^[22]对英国具有乳腺癌家族史的 408 例乳腺癌患者和 400 例健康对照使用含 227 876 个 SNP 位点的芯片进行第 1 阶段检测,根据病例组与对照组的等位基因变化频率的差异选择出 12 711 个 SNP,对 3 990 例乳腺癌患者和 3 916 例对照进行第 2 阶段验证,结合 2 个阶段的统计挑选

DOI: 10.3877/cma.j.issn.1674-0807.2019.04.009

基金项目:国家自然科学基金资助项目[重大国际(地区)合作与交流项目 81320108015]

作者单位:515041 汕头大学医学院长江学者实验室¹; 515041 汕头大学医学院附属肿瘤医院乳腺中心²; 361000 厦门大学附属翔安医院乳腺甲外科³; 523000 东莞市人民医院乳腺科⁴

通信作者:张国君 Email: guoj_zhang@yahoo.com

出 30 个 SNP ,并对 21 860 例浸润性乳腺癌患者、988 例 DCIS 患者和 22 578 例健康对照进行第 3 阶段验证。结果提示 5 个分别位于基因 FGFR2、TNRC9、MAP3K1、LSP1 及 8q24 上的 SNP(rs2981582、rs3803662、rs889312、rs381798、rs13281615) 与乳腺癌风险相关($P < 1 \times 10^{-7}$)。Garcia 等^[23]通过分析上述乳腺癌易感性位点与 13 527 例浸润性乳腺癌患者生存的相关性,发现 rs13281615 与乳腺癌预后显著相关($HR = 0.90$, $95\% CI: 0.83 \sim 0.97$, $P = 0.009$) ,该 SNP 位点位于 8q24.21 ,该位置目前未含已知的基因,但含有众多与前列腺癌和结直肠癌相关的独立风险位点。Sapkota 等^[24]对加拿大人群进行的乳腺癌预后 GWAS 也证实了 rs13281615 与生存显著相关,并提示其与 DFS 率($HR = 0.77$, $95\% CI: 0.65 \sim 0.90$, $P = 0.001$) 和 OS 率($HR = 0.76$, $95\% CI: 0.64 \sim 0.91$, $P = 0.003$) 的提高有相关性。这些研究不仅验证了已发现的乳腺癌易感基因,而且还意味着通过 GWAS 能够确定这些高风险位点与乳腺癌预后的联系。

Azzato 等^[25]使用了包含 528 252 个 SNP 位点的 Illumina 芯片对 1 145 例绝经后乳腺癌患者样本进行第 1 阶段检测,选择统计学差异最显著的 10 个 SNP 位点[差异最明显的 SNP(rs13124167) 是位于 4q31.23 上的 ARHGAP10 基因($P = 2.3 \times 10^{-7}$)] ,然后对 4 335 例乳腺癌患者样本进行第 2 阶段检测,结果未发现任何达 GWAS 显著性水准(5×10^{-8}) 的位点与乳腺癌生存相关。该研究提示寻找与乳腺癌生存显著相关($HR > 2$) 的 SNP 是十分困难的,并且,因单个 GWAS 样本量的限制,常难以检出变异频率较低的 SNP; 根据不同肿瘤类型和治疗,将研究样本分成不同的亚组,针对性更强,或许能找到与乳腺癌预后相关的 SNP。

Shu 等^[26]使用包含 613 031 个 SNP 的 Affymetrix 芯片,对来自上海的 1 950 例乳腺癌患者(包括 I ~ IV 期) 进行探索阶段的检测,筛选出 49 个 SNP 并在同一人群中的另外 4 160 例患者进行验证,发现 rs3784099 和 rs9934948 与乳腺癌病死率的上升明显相关($HR = 1.41$, $95\% CI: 1.18 \sim 1.68$, $P < 5.0 \times 10^{-6}$)。这一发现在欧洲人群中也得到了验证。rs3784099 位于 14 号染色体 RAD51L1 基因的 7 号内含子区, RAD51L1 是一种癌症易感基因,其编码的蛋白属于 RAD51 家族,经过同源重组途径影响 DNA 修复过程,过表达该基因会导致细胞周期延长和细胞凋亡^[27]。此外,该位点还与 SNCG 和 CTF1 基因的表达存在相关性,SNCG 基因通过激活 Rho 蛋白家族的小 GTP 酶(small GTPase) 和细胞外信号调节激酶来增加细胞的运动能力,可能影响乳腺癌和卵巢癌的转移^[28-29]。CTF1 基因作为转录因子能够通过划定染色质的界限来阻断沉默染色质的延长^[30]。这些基因都参与到恶性肿瘤的发生、发展过程中。rs9934948 则位于 16 号染色体的中间区域,距 PSMD7 基因 891 kb,研究表明该基因作为蛋白酶的组成部分,在肿瘤细胞中活性增强时可影响细胞的增殖和凋亡过程^[31]。上述生物学研究进一步证实了该研究在中国人群与欧洲人群所观察到的与病死率显著相关的位点可能与乳腺癌预后相关。然而,此位点在 CGEMS 研究^[16]中却无法得到验证,原因可能在于 CGEMS 的研究对象几乎均为绝经后女性,与上述其他研究的入组人群(年轻乳腺癌患

者) 不同。因此,该 SNP 与乳腺癌预后的关联性是否会因为月经状态而改变还有待验证。

Azzato 等^[32]选择了 3 990 例浸润性乳腺癌患者和 3 916 例对照(未患乳腺癌人群) 进行筛查,然后选择来自国际乳腺癌协会联盟的 15 项病例对照研究中的 14 096 例患者(来自欧洲、北美、澳大利亚) 进行验证,结果发现位于 15q13.1 的 OCA2 基因上的 rs4778137 与 ER 阴性乳腺癌患者的 OS 良好相关($P = 9.2 \times 10^{-5}$)。OCA2 基因与皮肤色素沉积、肤色及虹膜颜色有一定的相关性^[33]。OCA2 基因编码的 OCA2 蛋白通常不会出现在正常的乳腺组织中,但在乳腺浸润性导管癌的组织中可检测到大量 OCA2 蛋白。该蛋白能调节细胞内谷胱甘肽的代谢,从而影响细胞对化疗药物顺铂和多柔比星的敏感性,而多柔比星和其他蒽环类药物经常用于 ER 阴性乳腺癌的治疗^[34]。这表明 rs4778137 与乳腺癌患者的生存率上升相关,其原因可能是乳腺癌患者对蒽环类药物治疗的敏感度增加。上述研究表明: SNP 位点有可能通过不同机制来影响乳腺癌预后,比如不同类型肿瘤的发展、宿主对肿瘤的反应以及乳腺癌治疗药物的代谢途径等。通过 GWAS 对 SNP 的精细定位,能够促进功能研究的发展,从而进一步阐明肿瘤、宿主与治疗反应之间的复杂关系。

Rafiq 等^[35]在欧洲人群进行了 2 个阶段的 GWAS,第 1 阶段选择 536 例早发性乳腺癌患者样本(发病年龄 ≤ 40 岁) 进行检测,选择 35 个 SNP($P < 1 \times 10^{-4}$) 对来自同一人群、同年龄段的其他样本(1 516 例) 进行第 2 阶段检测,结果发现位于 5q 的 ARRDC3 基因上游 556 kb 的 rs421379 与早发性乳腺癌患者的 OS 存在相关性($HR = 1.61$, $95\% CI: 1.33 \sim 1.96$, $P = 9.5 \times 10^{-7}$)。然而,这一发现却无法在芬兰人群(805 例非早发性乳腺癌样本) 中得到重复验证^[36]。选择病死率较高、预后较差的乳腺癌患者进行检测,是确定与预后相关的 SNP 的一种有效策略,但要真正在临床上常规检测这些可能影响预后的突变位点,还需要分别进行辅助或新辅助治疗的相关随机临床试验来证实它们的预测价值。

四、结语

关于乳腺癌易感性的 GWAS 已经在全世界广泛开展,研究者也逐渐报道了不同人群的乳腺癌风险位点。尽管已有研究表明某些 SNP 与乳腺癌患者的生存相关,但这些研究仍存在一定的局限性,还需要更多的 GWAS 去寻找有预后价值的易感位点,并对其进行精细定位和开展机制研究,以及在前瞻性临床试验中进行验证。由此确定的 SNP 或许能够与其他预后因素一起为乳腺癌患者的预后预测提供有效信息,并为乳腺癌的预防和精准治疗提供理论依据及靶点。

参 考 文 献

- [1] Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012[J]. CA Cancer J Clin, 2015, 65(2): 87-108.
- [2] Kaplan HG, Malmgren JA, Atwood MK, et al. Effect of treatment and mammography detection on breast cancer survival over time: 1990 - 2007[J]. Cancer, 2015, 121(15): 2553-2561.
- [3] Dong G, Wang D, Liang X, et al. Factors related to survival rates for breast cancer patients[J]. Int J Clin Exp Med, 2014, 7(10): 3719-3724.

- [4] Corso G , Feroce I , Intra M , et al. Prognostic impact of germline mutations in inherited cancer syndromes [J]. *Future Oncol* , 2017 , 13 (24) : 2125-2127.
- [5] Fachal L , Dunning AM. From candidate gene studies to GWAS and post-GWAS analyses in breast cancer [J]. *Curr Opin Genet Dev* , 2015 , 30 : 32-41.
- [6] Maxwell KN , Nathanson KL. Common breast cancer risk variants in the post-COGS era: a comprehensive review [J]. *Breast Cancer Res* , 2013 , 15 (6) : 212.
- [7] Purcell S , Neale B , Todd-Brown K , et al. PLINK: a tool set for whole-genome association and population-based linkage analyses [J]. *Am J Hum Genet* , 2007 , 81 (3) : 559-575.
- [8] Edwards AO , Ritter R 3rd , Abel KJ , et al. Complement factor H polymorphism and age-related macular degeneration [J]. *Science* , 2005 , 308 (5720) : 421-424.
- [9] Fall T , Ingelsson E. Genome-wide association studies of obesity and metabolic syndrome [J]. *Mol Cell Endocrinol* , 2014 , 382 (1) : 740-757.
- [10] Chen X , Gustafsson S , Whittington T , et al. A genome-wide association study of IgM antibody against phosphorylcholine: shared genetics and phenotypic relationship to chronic lymphocytic leukemia [J]. *Hum Mol Genet* , 2018 , 27 (10) : 1809-1818.
- [11] Liu J , Wang L , Qian Y , et al. Association of 48 type 2 diabetes susceptibility loci with fasting plasma glucose and lipid levels in Chinese Hans [J]. *Diabetes Res Clin Pract* , 2018 , 139 : 114-121.
- [12] Forstner AJ , Hecker J , Hofmann A , et al. Identification of shared risk loci and pathways for bipolar disorder and schizophrenia [J]. *PLoS One* , 2017 , 12 (2) : e0171595.
- [13] Li Y , Wang DW , Chen Y , et al. Genome-wide association and functional studies identify SCML4 and THSD7A as novel susceptibility genes for coronary artery disease [J]. *Arterioscler Thromb Vasc Biol* , 2018 , 38 (4) : 964-975.
- [14] Visscher PM , Brown MA , McCarthy MI , et al. Five years of GWAS discovery [J]. *Am J Hum Genet* , 2012 , 90 (1) : 7-24.
- [15] MacArthur J , Bowler E , Cerezo M , et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog) [J]. *Nucleic Acids Res* , 2017 , 45 (D1) : D896-D901.
- [16] Hunter DJ , Kraft P , Jacobs KB , et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer [J]. *Nat Genet* , 2007 , 39 (7) : 870-874.
- [17] Stacey SN , Manolescu A , Sulem P , et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer [J]. *Nat Genet* , 2007 , 39 (7) : 865-869.
- [18] Turnbull C , Ahmed S , Morrison J , et al. Genome-wide association study identifies five new breast cancer susceptibility loci [J]. *Nat Genet* , 2010 , 42 (6) : 504-507.
- [19] Zheng W , Long J , Gao YT , et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1 [J]. *Nat Genet* , 2009 , 41 (3) : 324-328.
- [20] Long J , Cai Q , Sung H , et al. Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer [J]. *PLoS Genet* , 2012 , 8 (2) : e1002532.
- [21] Cai Q , Long J , Lu W , et al. Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium [J]. *Hum Mol Genet* , 2011 , 20 (24) : 4991-4999.
- [22] Easton DF , Pooley KA , Dunning AM , et al. Genome-wide association study identifies novel breast cancer susceptibility loci [J]. *Nature* , 2007 , 447 (7148) : 1087-1093.
- [23] Garcia-Closas M , Hall P , Nevanlinna H , et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics [J]. *PLoS Genet* , 2008 , 4 (4) : e1000054.
- [24] Sapkota Y , Yasui Y , Lai R , et al. Identification of a breast cancer susceptibility locus at 4q31.22 using a genome-wide association study paradigm [J]. *PLoS One* , 2013 , 8 (5) : e62550.
- [25] Azzato EM , Pharoah PD , Harrington P , et al. A genome-wide association study of prognosis in breast cancer [J]. *Cancer Epidemiol Biomarkers Prev* , 2010 , 19 (4) : 1140-1143.
- [26] Shu XO , Long J , Lu W , et al. Novel genetic markers of breast cancer survival identified by a genome-wide association study [J]. *Cancer Res* , 2012 , 72 (5) : 1182-1189.
- [27] Shailani A , Kaur RP , Munshi A. A comprehensive analysis of BRCA2 gene: focus on mechanistic aspects of its functions , spectrum of deleterious mutations , and therapeutic strategies targeting BRCA2-deficient tumors [J]. *Med Oncol* , 2018 , 35 (3) : 18.
- [28] He K , Wang P. Unregulated long non-coding RNA-AK058003 promotes the proliferation , invasion and metastasis of breast cancer by regulating the expression levels of the gamma-synuclein gene [J]. *Exp Ther Med* , 2015 , 9 (5) : 1727-1732.
- [29] Min L , Zhang C , Ma R , et al. Overexpression of synuclein-gamma predicts lack of benefit from radiotherapy for breast cancer patients [J]. *BMC Cancer* , 2016 , 16 : 717.
- [30] Lutz SZ , Franck O , Bohm A , et al. Common genetic variation in the human CTF1 locus , encoding cardiotrophin-1 , determines insulin sensitivity [J]. *PLoS One* , 2014 , 9 (7) : e100391.
- [31] Almond JB , Cohen GM. The proteasome: a novel target for cancer chemotherapy [J]. *Leukemia* , 2002 , 16 (4) : 433-443.
- [32] Azzato EM , Tyrer J , Fasching PA , et al. Association between a germline OCA2 polymorphism at chromosome 15q13.1 and estrogen receptor-negative breast cancer survival [J]. *J Natl Cancer Inst* , 2010 , 102 (9) : 650-662.
- [33] Crawford NG , Kelly DE. Loci associated with skin pigmentation identified in African populations [J]. *Science* , 2017 , 358 (6365) : pii: eaan8433.
- [34] Li XP , Lan JY , Liu DQ , et al. OCA2 rs4778137 polymorphism predicts survival of breast cancer patients receiving neoadjuvant chemotherapy [J]. *Gene* , 2018 , 651 : 161-165.
- [35] Rafiq S , Tapper W , Collins A , et al. Identification of inherited genetic variations influencing prognosis in early-onset breast cancer [J]. *Cancer Res* , 2013 , 73 (6) : 1883-1891.
- [36] Fagerholm R , Hofstetter B , Tommiska J , et al. NAD(P) H: quinone oxidoreductase 1 NQO1* 2 genotype (P187S) is a strong prognostic and predictive factor in breast cancer [J]. *Nat Genet* , 2008 , 40 (7) : 844-853.

(收稿日期: 2017-06-02)

叶娴 郭翠萍 张国君. 基于全基因组关联研究的乳腺癌预后预测研究现状 [J/CD]. 中华乳腺病杂志(电子版) , 2019 , 13 (4) : 239-241.